

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

520.40577X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

#4
JC997 U.S. PTO
09/942615
08/31/01

Applicant(s): Junichi KIMURA, et al
Serial No.:
Filed: August 31, 2001
Title: MEDIA DISTRIBUTION SYSTEMS AND MULTI-MEDIA
CONVERSION SERVER
Group:

LETTER CLAIMING RIGHT OF PRIORITY

Honorable Commissioner of
Patents and Trademarks
Washington, D.C. 20231

August 31, 2001

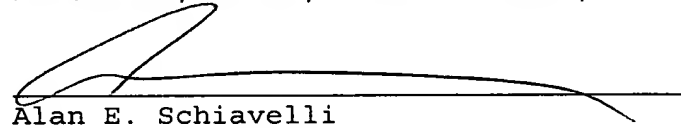
Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the
applicant(s) hereby claim(s) the right of priority based on Japanese
Patent Application No.(s) 2001-102922 filed April 2, 2001.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP


Alan E. Schiavelli
Registration No. 32,087

AES/nac
Attachment
(703) 312-6600

日 本 国 特 許 庁
JAPAN PATENT OFFICE

JC997 U.S. PTO
09/942615
08/31/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年 4月 2日

出 願 番 号

Application Number:

特願2001-102922

出 願 人

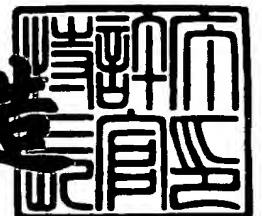
Applicant(s):

株式会社日立製作所

2001年 8月10日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3072216

【書類名】 特許願

【整理番号】 NT01P0071

【提出日】 平成13年 4月 2日

【あて先】 特許庁長官 殿

【国際特許分類】 H04L 12/58

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所 中央研究所内

【氏名】 木村 淳一

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所 中央研究所内

【氏名】 鈴木 芳典

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所 中央研究所内

【氏名】 永松 健司

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100068504

【弁理士】

【氏名又は名称】 小川 勝男

【電話番号】 03-3661-0071

【選任した代理人】

【識別番号】 100086656

【弁理士】

【氏名又は名称】 田中 恭助

【電話番号】 03-3661-0071

【選任した代理人】

【識別番号】 100094352

【弁理士】

【氏名又は名称】 佐々木 孝

【電話番号】 03-3661-0071

【手数料の表示】

【予納台帳番号】 081423

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 メディア配信システム及びマルチメディア変換サーバ

【特許請求の範囲】

【請求項 1】

第 1 の端末から送信されたメディア情報を第 2 の端末に配信するサーバとを有するメディア配信システムにおいて、上記サーバが上記第 2 の端末のメディア再生能力を取得する手段と、上記メディア情報を上記取得したメディア再生能力に応じた出力メディア情報に変換する手段とを具備するメディア配信システム。

【請求項 2】

第 1 の端末から送信されたメディア情報を受信する手段と、受信した上記メディア情報の宛先を取得する手段と、上記宛先である第 2 の端末のメディア再生能力を取得する手段と、上記メディア情報を上記第 2 の端末のメディア再生能力に応じた出力メディア情報に変換する手段と、上記第 2 の端末に対して上記出力メディア情報を送信する手段とを具備するマルチメディア変換サーバ。

【請求項 3】

第 1 の端末から第 2 の端末に宛てた文字情報を受信する手段と、上記文字情報を音声信号に変換する音声信号変換手段と、上記音声信号に対応した映像信号を生成する映像信号生成手段と、上記音声信号を第 2 の端末が受信再生できるフォーマットの 1 つで圧縮符号化する音声信号圧縮手段と、上記映像信号を第 2 の端末が受信再生できるフォーマットの 1 つで圧縮符号化する映像信号圧縮手段と、上記文字情報に圧縮した音声符号と圧縮した映像符号を付加し上記第 2 の端末に送信する手段とを具備するマルチメディア変換サーバ。

【請求項 4】

請求項 3 記載のマルチメディア変換サーバであって、更に上記第 2 の端末が受信再生できるフォーマット情報を入手する手段を有し、上記音声信号圧縮手段及び映像信号圧縮手段が、上記音フォーマット情報を用いて圧縮を行うように構成されたマルチメディア変換サーバ。

【請求項 5】

請求項 3 記載の変換サーバにおいて、更に、上記第 1 の端末に対して複数の変

換する音声の種類と複数の生成する映像の種類を呈示し、これら音声及び映像のうちそれぞれ1種類の選択指示する手段を有し、上記音声信号変換手段は上記選択した音声選択情報と映像選択情報とを文字情報に付された受信信号の選択された音声選択情報の内容に従った音声信号に変換するように構成され、上記映像信号生成手段は上記選択された映像信号を合成するように構成されたことを特徴とするマルチメディア変換サーバ。

【請求項 6】

請求項5のマルチメディア変換サーバと通信を行うマルチメディア端末において、文字を入力編集する手段と、上記変換する音声の種類を呈示し、選択された音声選択情報を生成する手段と、上記映像の種類を呈示し、選択れた映像選択情報をする機能と、入力した文字情報と合成音声選択情報と、合成映像選択情報を送信するマルチメディア端末。

【請求項 7】

第1の端末から第2の端末に宛てた映像情報を受信する手段と、上記第2の端末が受信再生できる映像符号フォーマット情報を入手する手段と、上記受信した映像情報の映像符号フォーマットを上記第2の端末が受信再生できる映像符号フォーマットと比較する手段と、上記比較した結果一致する受信した映像情報を第2の端末が受信再生できる映像符号フォーマットがなければ第2の端末が受信再生できる映像符号フォーマットの1つを選択し、上記入力した映像情報を選択した映像符号フォーマットへ変換する手段と、上記変換した映像情報を端末2に送信する手段とを具備するマルチメディア変換サーバ。

【請求項 8】

第1の端末から第2の端末に宛てた映像情報を受信する手段と、上記第2の端末が受信再生できる画面サイズ情報を入手する手段と、受信した映像情報の画面サイズを第2の端末が受信再生できる画面サイズ情報と比較する手段と、比較した結果受信した映像情報の画面サイズを上記第2の端末が受信再生できる画面サイズより大きければ入力した映像情報を第2の端末が受信再生できる画面サイズへ変換する手段と、変換した映像情報を端末2へ宛てて送信する手段とを具備するマルチメディア変換サーバ。

【請求項 9】

請求項 1 記載のマルチメディア変換サーバにおいて、入力メディア情報の種類と出力メディア情報の種類の組み合わせにより定められる変換料金を送信者に対して課金することを特徴とするマルチメディア変換サービス。

【請求項 1 0】

請求項 3、4、6 又は 7 に記載の変換サーバにおいて、受信した文字情報から音声情報あるいは映像情報への変換をした場合、変換をしない場合に比べ高い料金を送信者に対して課金することを特徴とするマルチメディア通信サービス。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、メディア配信システム及びマルチメディア変換サーバ、更に詳しく言えば、映像及び音声情報を含む情報を送受信する通信システムに用いる携帯マルチメディア端末及び携帯マルチメディア端末間の通信データを中継するマルチメディアサーバに関するものである。

【0 0 0 2】

【従来の技術】

映像信号（動画像）及び音声あるいは音楽信号は国際標準規格 ISO/IEC 14496（MPEG-4）等を用いることにより、数十kbit/秒（以下bpsと略する）程度に圧縮して伝送することができる。また、一定時間の映像・音声信号をMPEG-4を用いて圧縮し、得られた符号データを 1 つあるいは映像、音声の 2 つのファイルとして電子メールデータ（テキスト情報）とあわせて送信することができる。

【0 0 0 3】

従来のマルチメディア端末による、映像・音声ファイルの送受信は、送信端末で映像・音声圧縮して伝送路を介して、配信サーバ（例えばメールサーバ）に転送する。配信サーバは受信したデータの宛先に該当する受信端末に、メールを転送する。あるいは、配信サーバは受信端末が配信サーバに接続することを監視し、接続したことを確認した時に、受信端末にメールが到来している旨、あるいはメール自体を受信端末に転送する。

【0004】

上記送信端末は、送信すべき文字入力情報（例えば押下キー情報）、映像信号、音声信号を入力し、文字入力情報は編集装置で解読され、文字コードとなりメモリにテキスト情報として格納される。上記映像信号は、映像符号に変換され、メモリに格納される。上記音声信号は、音声符号に変換され、メモリに格納される。送信端末ユーザの指示により、送信端末は、配信サーバ呼び出し、伝送路を確立する。次に、上記メモリに格納されたテキスト情報（メールの宛先、本文等）、映像符号、音声符号が、読み出され、確立した伝送路を介して、サーバへ送信される。

【0005】

上記伝送路上での送信情報は、宛先、テキスト情報、音声情報、映像情報が一定のフォーマットで伝送される。送信端末からのデータ（以下メールデータ）を受信した配信サーバは、入力された情報をバッファに格納する。このとき、必要に応じて課金制御部で、配信サーバが受信した情報量に応じた料金を送信者に対して課金するための記録をとる。その後、バッファに格納されたメールデータから、その宛先を解読して、宛先に該当する受信端末を呼出す。配信サーバと受信端末との伝送路が確立した時点で、バッファに格納されているメール情報（テキスト情報、音声情報、映像情報）を読み出し、受信端末にメールデータを送信する。

【0006】

上記受信端末は配信サーバからの呼出しを受けると、配信サーバとの間に伝送路を確立し、配信サーバから伝送されたメール情報を、メモリに格納する。受信端末のユーザは、受信したメール情報を選択し、テキスト表示処理をして表示デバイス上に表示して読む。また、必要に応じて映像符号、音声符号を読み出し、映像信号、音声信号を再生する。

【0007】

また、上述のマルチメディア配信システムでは、映像情報符号を生成するために画像入力カメラ及び映像エンコーダを実装する必要がある上、コスト高になる上、多くの電力を必要とするため送信端末を駆動する電池の寿命が短くなり、より

大容量の電池を搭載することにより端末のサイズが大きくなり携帯性が損なわれる問題があり、さらに、送信端末と受信端末の間で、同一の映像情報符号処理アルゴリズムを実装する必要が生じ、通信相手選択の範囲が狭められてしまう問題がある。この問題を解決するため、他の従来例として、特開平 6 - 1 6 2 1 6 7 号公報に開示されているように、受信端末にて、受信文字情報に合せて音声、画像を合成し、その際に使用するパラメータを送信端末で指定する方法が知られている。

【 0 0 0 8 】

【発明が解決しようとする課題】

上記他の従来例では、送信端末及び配信サーバでの情報処理量及び伝送容量は、軽減されるが、受信端末で合成処理を行うため、多くの処理能力が必要となり、コストが高価になる上、多くの電力を必要とするため送信端末を駆動する電池の寿命が短くなり、より大容量の電池を搭載することにより端末のサイズが大きくなり携帯性が損なわれる点が考慮されていない。さらに、送信端末で、受信端末の合成アルゴリズムのパラメータを事前に知る必要があり、合成アルゴリズムのメンテナンス性、拡張性に欠ける点が考慮されていない。

【 0 0 0 9 】

従って、本発明の第 1 の目的は、送信端末と受信端末の間で、同一のメディア情報符号処理アルゴリズムが異なる場合にも配信できるマルチメディア配信システム及びそれに用いるサーバを実現することである。

【 0 0 1 0 】

本発明の他の目的は、第 1 の目的を達成すると同時に送信端末及び受信端末のデータ処理量を軽減し、消費電力、使用コストを軽減できるのマルチメディア配信サーバを実現することである。

【 0 0 1 1 】

【課題を解決するための手段】

上記目的を達成するため、本発明は、送信、受信端末間のマルチメディア通信データを中継するサーバを介してメディア情報（テキスト、映像及び音声情報）を伝送・受信する配信システムにおいて、上記サーバに上記受信端末のメディア再

生能力を取得する手段と、上記送信端末からのメディア情報を上記取得したメディア再生能力に応じた出力メディア情報に変換する手段と設けて構成する。以下上記構成のサーバをマルチメディア変換サーバと呼ぶ。

【 0 0 1 2 】

そのため、本発明のマルチメディア変換サーバは第 1 の端末（送信端末）から送信されたメディア情報を受信する受信手段、受信した上記メディア情報の宛先を取得する手段、その宛先である第 2 の端末（受信端末）のメディア再生能力を取得する手段、上記メディア情報を上記受信端末のメディア再生能力に応じた出力メディア情報に変換する変換手段、上記受信端末に対して上記出力メディア情報を送信する出力手段を設けて構成される。

【 0 0 1 3 】

本発明のマルチメディア変換サーバの好ましい実施形態として、上記受信手段が受信するメディア情報が文字情報で、上記メディア再生能力はフォーマット情報であり、上記変換手段は上記文字情報を音声信号に変換する手段、生成した音声に対応した映像信号を生成する手段、生成した音声信号を第 2 の端末が受信再生できるフォーマットの 1 つにて圧縮符号化する手段、生成した映像信号を第 2 の端末が受信再生できるフォーマットの 1 つにて圧縮符号化する手段とをもち、上記出力手段は、上記文字情報に圧縮した上記音声符号と圧縮した上記映像符号を付加し受信端末へ宛て送信する手段とをもち。

【 0 0 1 4 】

本発明では、送信端末は受信端末の画像合成アルゴリズムのメディア再生能力等を知らなくても通信が可能となる。またテキスト情報を基に音声、映像情報を合成して生成することにより、送信端末及び受信端末の処理量を低減し、携帯端末の小型化、端末電池の長寿命化を実現できる。

【 0 0 1 5 】

本発明の上述及び他の特徴及び効果は以下の発明の実施の形態によって、更に詳しく説明する。なお、今後の説明において、音声の各音に対応する情報を音素片情報、音素片を組み合わせた一連の情報を音声情報、動画像を構成する各画面を画像あるいはフレーム、画像あるいはフレームを組み合わせた一連の情報を映像

情報と呼ぶ。

【 0 0 1 6 】

【発明の実施の形態】

図 1 は本発明によるマルチメディア配信システムの第 1 の実施形態を示す構成ブロック図である。本実施形態はマルチメディア端末による、映像・音声ファイルの送受信を、送信端末は、受信端末の処理能力を不知の状態の情報で伝送できるものである。

【 0 0 1 7 】

本システムは送信端末 1 0 0 から送信されたメディア情報を受信端末 5 に配信するサーバとを有するメディア配信システムにおいて、上記サーバが端末データベースサーバ 1 0 7 を用いて上記受信端末 5 のメディア再生能力を取得する手段と、上記メディア情報を上記取得したメディア再生能力に応じた出力メディア情報に変換する音声・映像合成サーバ 1 0 3 で構成される。

【 0 0 1 8 】

送信端末 1 0 0 は配信サーバ 1 0 1 に対して伝送路 2 を介して、受信端末の識別情報(端末 ID) 2 1 0 1、テキスト情報と予め定められた映像及び音声のうちのそれぞれ 1 つを選択する選択信号のみを送信する。配信サーバ 1 0 1 は送信先である受信端末 5 の処理能力を、端末データベースサーバ 1 0 7 に対して、受信端末 5 の識別情報 2 1 0 1 を通知し、受信端末 5 の処理能力を問い合わせる。

【 0 0 1 9 】

端末データベースサーバ 1 0 7 は受信端末 5 の視聴可能な音声符号フォーマット、映像符号フォーマット、画面サイズ等の音声映像再生能力情報 2 1 0 2 を配信サーバ 1 0 1 に通知し、配信サーバ 1 0 1 は音声映像再生能力情報 2 1 0 2 を基に音声及び映像符号化方式を決定する。配信サーバ 1 0 1 は受信したテキスト情報 1 0 2 と、映像選択信号 1 0 6、音声選択信号 1 0 5、音声映像符号化方式 1 0 8 を音声・映像合成サーバ 1 0 3 に送信する。

【 0 0 2 0 】

音声・映像合成サーバ 1 0 3 では、テキスト情報 1 0 2 を基に、テキストに記された内容を基に音声信号、映像信号を合成・符号化し、得られた音声・映像符

号 1 0 4 を配信サーバ 1 0 1 に返す。配信サーバ 1 0 1 では、送信端末 1 0 0 から送信されたテキスト情報と、音声・映像合成サーバから得た音声/映像符号 1 0 4 を伝送路 4 を介して受信端末 5 に伝送する。受信端末 5 は受信した信号を解読し、テキスト情報、映像信号、音声信号をそれぞれ表示・再生する。

【 0 0 2 1 】

図 2 は、図 1 の音声映像再生能力情報 2 1 0 2 を取得する手順を示すフロー図である。配信サーバ 1 0 1 は端末 DB サーバ 1 0 7 に対して、端末能力問合せ要求信号と端末能力取得要求を送付する。端末の識別情報(端末 ID)はメールアドレス、電話番号、機器番号、機器型番号等である。配信サーバ 1 0 1 は、端末能力問合せ要求信号 2101 が受け付けられた旨を示す了解応答を受信した後、端末 ID を送信し、端末 DB サーバ 1 0 7 は該当する音声映像再生能力情報 2 1 0 2 を返送する。配信サーバ 1 0 1 は音声映像再生能力情報 2 1 0 2 を受信した後、終了要求を通知し、音声映像再生能力情報の受信処理を終了する。

【 0 0 2 2 】

図 3 は、端末 DB サーバ 1 0 7 における音声映像再生能力情報管理の一例を示す。端末 DB サーバ 1 0 7 では、図 3 に示すような端末 ID とその ID に対応する音声映像再生能力情報をセットにしたテーブルをもつ。配信サーバ 1 0 1 から音声映像再生能力情報取得の要求が来ると、これに付随して通知される端末 ID を用いて、図 3 のテーブルを検索し、得られた音声映像再生能力情報 2 1 0 2 を返送する。

【 0 0 2 3 】

図 4 は配信サーバ 1 0 1 へ返信する端末能力送信フォーマットと音声映像再生能力情報(端末能力)を示す。端末能力送信フォーマット 5 0 5 0 は、識別フィールド、端末 ID フィールド、端末能力フィールド、検証フィールドの 4 部分から構成される。識別フィールドはこれに引き続くデータで端末能力が送信されることを示す符号である。端末 ID フィールドは、配信サーバ 1 0 1 から要求された端末 ID を返送する。配信サーバ 1 0 1 で、端末 ID フィールドの情報と、要求した端末 ID とを比較することにより、受信したデータの正当性を確認する。端末能力フィールドは図 4 の引き出し先により示されるように音声、映像それぞ

れに関して端末の能力を示すデータ（音声映像再生能力情報 5 0 5 1）である。検証フィールドは識別フィールド、端末 I D フィールド、端末能力フィールドのデータ（ビット、バイト）等に伝送エラーがないことを確認するための情報であり、例えば、パリティや、CRC 符号等が該当する。また、さらに、誤り訂正符号（例えばリードソロモン符号、BCH 符号等）を用いて、軽微な伝送エラーであれば受信側にて修正させる機構を設けてもよい。

【 0 0 2 4 】

図 4 の下部に音声映像再生能力情報 5 0 5 1 の詳細を示す。音声能力情報、映像能力情報ともに、方式フラグと能力値の 2 つの部分からなる。方式フラグは候補となる複数の方式、オプション等にそれぞれフラグを設け、それぞれの方式をサポートしていれば TRUE（真）、サポートしていなければ FALSE（偽）をセットする。図 4 では、音声符号化方式として A、B、C の 3 方式の候補、映像符号化では P、Q、R、S の 4 候補があり、図の例では音声は方式 A のみ、映像は方式 Q 以外をサポートしていることを示す（1 = TRUE）。能力値は方式フラグに示す方式に付随した数値的限界を示すものであり、例えば、ビットレート（図の "B-rate"、"B-rate2"）、音声処理における音声サンプリングレート（図の "S-rate"）、映像処理における最大画像サイズ（図の "size"）、フレームレート（図の "F-rate"）等が例として挙げられる。能力値は、ビットレート、フレームレート等のように数値で表すもの、サンプリングレートのように、予め設定された数値に対し真偽を示す値を記すもの、画像サイズのように複数のスカラ値の組み合わせにより示すもの等がある。また、これらを符号化する方法、予め定められた複数の「値の範囲」の中から選択する方法等もある。また、方式フラグ、能力値ともに、「拡張フラグ」を設け、このフラグが真の時には、新たなフィールドが追加される構造にすることにより、将来の方式数の増加等にも互換性を保ちながら拡張することができる。さらに、音声、映像の能力以外にもテキストや、グラフィックス、通信方式、高音質オーディオ等の能力も同様な記述法にて記述することができる。

【 0 0 2 5 】

図 5 は配信サーバ 1 0 1 における、音声映像再生能力情報 5 0 5 1 の音声能力

の処理フロー図である。配信サーバ101は受信した音声映像再生能力情報5051を解読しながら、まず、判定部5101で方式Aがサポートされているか、すなわちフラグが1か否かを判定し、方式Aがサポートされていれば、関連する能力値、すなわちサンプリングレート5102、ビットレート5103をデータから取得・設定し、正常終了する。方式Aがサポートされていない場合は方式B、方式Bがサポートされていない場合は方式Cを調べる。いずれかの方式がサポートされていれば関連する能力値を取得し正常終了する。

【0026】

図では、方式Bではサンプリングレート、ビットレートは固定であるため能力値取得不要、方式Cはビットレートのみ可変であるため能力値取得することを仮定している（方式Aはサンプリングレート、ビットレートはいずれも選択可能を仮定）。方式A、B、Cのいずれもサポートしていない場合は、エラーとし、該当方式がない旨を送信端末100へ通知する。なお、以上の説明では、方式の判定はA→B→Cの順で優先順位をつけて判断しているが、これを可変とする、あるいはハードウェアの稼動状況に応じて可変としてもよい。

【0027】

図6は上記方式選択に優先順位を設けた選択法による処理フロー図である。同図において、まず、配列：優先順位テーブル[i]に希望する選択方式の順iに0から、選択方式を識別する情報（例えば方式番号等）を記述しておく。このとき、全選択方式数を「候補数」とする。まず、変数iを用いて、優先順位テーブルに記載された選択方式順に「選択方式候補」を選ぶ。また「選択方式候補」の方式に対応する「方式フラグ」を受信した配列：受信方式フラグ[]から選択する。この「方式フラグ」が1（真）か否かを調べ、真であれば、この時の「選択方式候補」を「選択方式」として採用して、以下、選択方式に応じた能力値を設定し、正常終了する。一方「方式フラグ」が0（偽）の時には、変数iを歩進した後、「候補数」と比較し、まだ、候補が残っていれば、再び、「選択方式候補」を選ぶステップへ戻る。そして、次の優先順位の方式の検査を行う。変数iと「候補数」の比較において、iが「候補数」と同じ、すなわち0から「候補数-1」までの「候補数」個の候補を検査し終わってしまった場合、該当候補なしとして

エラー終了とする。

【 0 0 2 8 】

図 6 の方法では、検査開始の前に、優先順位テーブルを設定すればよいため、随時優先順位を変更できる。また、優先順位テーブルにある方式を登録しないことにより、端末がその方式をサポートしていた（受信方式フラグ[]の対応フラグが真）としても、この方式を選択しないことができる。

【 0 0 2 9 】

図 7 は本発明の配信システムに使用される送信端末 1 0 0 及び受信端末 5 に相当するマルチメディア端末 1 0 0 0 の構成図である。説明を簡単にするため、送信機能のみを抽出した端末 1 0 0 と受信機能のみを抽出した端末 5 に分けて以下説明を行う。

【 0 0 3 0 】

図 8 は送信端末 1 0 0 の構成図であり、図 7 のマルチメディア端末 1 0 0 0 の送信機能のみを抽出したものである。送信端末 1 0 0 では、入力デバイス 1 1 から入力された文字入力情報 1 2 は、編集装置 1 3 で解読され文字コード 1 4 となり、テキスト情報（宛先情報、テキスト情報）としてメモリ 1 5 格納される。合わせて、受信側に送る合成映像信号、合成音声信号の種別を選択する選択部 1 1 0 より音声選択信号 1 1 1、映像選択信号 1 1 2 が選択され、メモリ 1 5 に格納される。送信時には、通信インターフェース（IF） 1 7 を介して、配信サーバ 1 0 1 との伝送路 2 を確立した後に、図 9 で示すような、宛先情報 5 0、音声・映像選択情報 1 1 5、テキスト情報 5 1 を配信サーバ 1 0 1 へ送信する。

【 0 0 3 1 】

図 1 0 は、合成音声・合成映像選択部 1 1 0 における音声・映像選択のための画面例である。選択のための情報はマルチメディア端末 1 0 0 0 の表示デバイス 6 6 上に表示され、表示するデータは、音声映像合成サーバ 1 0 3 より、配信サーバ 1 0 1 経由で受信しており、メモリ 1 5 上に格納されている。図 1 0 は 3 つの顔画像 1 0 0 2、1 0 0 3、1 0 0 4 から 1 つの顔画像を、また、3 種の音声 1 0 0 8、1 0 0 9、1 0 1 0 から 1 つの音声を選択する画面であり、顔画像はそれぞれボタン 1 0 0 5、1 0 0 6、1 0 0 7 で、音声はそれぞれボタン 1 0 1

1, 1 0 1 2, 1 0 1 3で選択する。図では画像 1（左端）及び音声 2（中央）を選択した様子を示している。この場合、図 9 の選択信号 1 1 5 としては画像 = 1、音声 = 2 を示す信号が伝送される。

【 0 0 3 2 】

図 1 1 は本発明によるマルチメディア変換サーバを構成する配信サーバの一実施形態の構成図である。配信サーバ 1 0 1 が従来知られている配信サーバと異なる点は音声・映像合成サーバ 1 0 3 と通信するための信号線 1 0 2、1 0 5、1 0 6、1 0 4 及び端末データベースサーバ 1 0 7 との通信するための信号線 1 0 8、2 1 0 1、2 1 0 2 が付加されている点である。

【 0 0 3 3 】

配信サーバ 1 0 1 の動作は 4 つのフェーズから構成される。第 1 のフェーズは送信端末 1 0 1 からのデータ（以下メールデータ）の受信であり、伝送路 2 から、通信 IF 4 1 を介して入力された情報 4 2 をバッファ 4 5 に格納する。このとき、必要に応じて課金制御部 4 3 で、配信サーバが受信した情報量及び音声・画像合成機能の使用／不使用、音声・画像合成する選択番号に応じた料金を送信者に対して課金するための記録をとる。例えば、音声・画像合成機能を使用する場合は使用しない場合の料金（A）よりも高額な料金（B）がとられ、差額（B-A）は音声・画像合成サーバの運営に費やされる。また、ある特定の画像を選択した場合にはさらに高額な料金（C）が課せられ、差額（C-B）は使用した画像の権利所有者に渡される。

【 0 0 3 4 】

第 2 及び第 3 のフェーズは音声・画像合成の機能を使う場合にのみ存在する。音声・画像合成の機能を使用するか否かは、図 9 における選択情報 1 1 5 が存在するか否か、あるいは選択情報 1 1 5 の内容が有効な情報を示しているか、「選択しない」ことを示す情報を示しているかにより判定される。また、常時フェーズ 2、3 が存在するよう端末とサーバの間で取り決めを行っていてもよい。また、別信号で通知してもよい。

【 0 0 3 5 】

第 2 のフェーズにおいては、配信サーバ 1 0 1 のコントロール部 2 1 0 3 は受

信したメールアドレスから宛先情報 2100 を抽出し、端末データベースサーバ 107 に対して受信端末の識別情報 2101 を送信し、受信端末 5 の音声映像再生能力情報 2102 を得る。コントロール部 2103 は受信端末 5 の再生能力に応じた音声符号化方式及び映像符号化方式を決定し、音声映像符号化方式 108 として音声・映像合成サーバ 103 へ通知する。

【0036】

第3のフェーズにおいては、配信サーバ 101 は受信したメールアドレスのコピーを音声・映像合成サーバ 103 に信号線 102 を介して送信する。音声・映像サーバ 103 で音声・映像を合成した結果の符号は信号線 104 を介して受信され、バッファ 45 に格納される。

【0037】

第4のフェーズは、第3のフェーズ（第3のフェーズが存在しない場合は第1のフェーズ）が終了した後の任意の時刻から開始される。第4のフェーズにおいては、通信制御部 47 がバッファに格納されたメールアドレス 46 を読出し、その宛先を解読する。そして、通信IF 49 に指示をして、宛先に該当する端末、すなわち受信端末 5 を呼び出す。受信端末 5 との伝送路 5 が確立した時点で、バッファ 45 に格納されているメール情報のテキスト情報及びもし存在すれば音声・映像合成符号を読出し、通信IF 49、伝送路 4 を介して、受信端末 5 にメールアドレスを送信する。

【0038】

図12は図6の音声・映像合成サーバ 103 の一実施形態の構成図である。図12の動作を説明する前に、図13及び図14を用いて音声・映像合成の原理を説明する。図13において「お願いします。」とのテキストを音声及び映像に変換する場合、まず、テキストを解析し、音情報「0 NE GA I SHI MA SU」に変換する。このとき、各音の継続時間、アクセントの位置等を決定する。変換した各音素片（例えば、「0」や「NE」）に対応する音声波形データを順次並べて行くことにより、入力したテキストに対応する音声を合成する。

【0039】

一方、画像合成では各音素片の種類に対応する画像を用意しておき、各音素片

の継続時間だけ、対応する画像を表示する。画像の種類としては、例えば、図 14 に示すように、7つのフレームを用意し、音に対応する画像を表示する。

フレーム 0 (第 14 図左端) 無声区間及び、ん、ま行、ば行、ぱ行の前半

フレーム 1 あ段 (あかさたなはまやらわがざだば) の音

フレーム 2 い段の音

フレーム 3 う段の音

フレーム 4 え段の音

フレーム 5 お段の音

フレーム 6 まばたき用

上記音情報「0 NE GA I SHI MA SU」の場合は、図 13 にも示したように、フレーム番号が 5→4→1→2→2→0→1→3 となるように画像を表示する。音声開始前、終了後、及び途中の無音区間はフレーム 0 を表示しておき、適宜 (例えば 2 秒間に 0.1 秒程度の比率にて) フレーム 6 を挿入することにより、まばたきをしているように見え、より自然な感じをユーザに与えることができる。

【0040】

図 12 に戻って、音声・映像合成サーバ 103 の動作を説明する。まず、音素片データベース 132 には各音に対応する音素片の波形データが格納されており、選択する声の種類 105 と音データ 133、必要であれば発生音の前後の音列、アクセント等の情報を与え、一意的に波形情報 134 を取り出す。また、画像データベース 128 には、図 14 に示したような複数のフレームが格納されており、選択する画像の種類 106 と音情報から得られる選択フレーム番号 126 が与えられれば、一意的にフレーム 127 が得られる。

【0041】

合成時には、テキスト情報 102 が音声変換部 120 に入力される。音声変換部 120 ではテキスト情報 102 から音への変換を行い、音データと各音の継続時間を決定する。変換された音データ 133 が音声データベース 132 に入力される。音声データベース 132 では、配信サーバ 101 から指定される音声選択信号 105 と、音データ 133 より、音声波形データ 134 を音声変換部 120 に出力する。音声変換部 120 では入力された音声波形データを上記継続時間だ

け音声出力波形信号 1 2 1 に出力する。出力された波形信号 1 2 1 は、そのままデジタル-アナログ変換を行えば実際の音（音声）となるが、音声・映像合成サーバ 1 0 3 においては、デジタル信号のまま音声エンコーダ 1 2 2 に入力し、音声映像符号化方式 1 0 8 の示す符号化方式で圧縮して、音声符号データ 1 2 3 を得る。

【 0 0 4 2 】

一方、音声変換部 1 2 0 は音データとその音の継続時間情報をフレーム選択部 1 2 5 に入力する。フレーム選択部 1 2 5 では音情報から表示するフレーム番号 1 2 6 を決定し、画像データベース 1 2 8 に入力する。画像データベース 1 2 8 では配信サーバ 1 0 1 から指定される画像選択信号 1 0 6 と、フレーム番号 1 2 6 より、表示フレームデータ 1 2 7 を出力する。フレーム選択部 1 2 5 は画像データベース 1 2 8 より入力された表示フレームデータ 1 2 7 を保持し、該当する音声信号 1 2 1 と同期するように、指定された継続時間の間、フレームデータ 1 2 9 を出力する。フレームデータ 1 2 9 は、表示フォーマットを変換しテレビジョン等で見ると口が動く動画像として見えるが、音声・映像合成サーバ 1 0 3 においては、デジタル信号のまま映像エンコーダ 1 3 0 に入力し、音声映像符号化方式 1 0 8 の示す映像符号化方式で圧縮し、映像符号データ 1 3 1 を得る。音声符号データ 1 2 3 と映像符号データ 1 3 1 はそれぞれが同期するように多重部 1 3 5 で 1 つの信号に多重化され、音声・映像符号データ 1 0 4 として配信サーバ 1 0 1 に戻される。

【 0 0 4 3 】

図 1 5 は本発明によるマルチメディア配信システムの第 2 の実施形態の構成図である。

第 1 の実形態と異なる点は、音声・映像合成処理を受信端末にて行う点、すなわち受信者が合成する音声・映像を選択する点である。送信端末 1 5 7 は図 8 の送信端末 1 0 0 とほぼ同じ構成であるが、合成音声・合成映像選択部がない、すなわち、テキスト情報のみを送信する端末である。送信されたテキスト情報は配信サーバ 3 を経由して、受信端末 1 5 0 に届く。

【 0 0 4 4 】

受信端末150は受信したテキスト情報を閲覧する前に、画像データベースサーバ152及び音素片データベースサーバ155に接続し、それぞれに対して、希望する画像選択信号151、音声選択信号154を送信し、該当するフレームデータセット153及び音素片波形セット156を得る。フレームデータセットは、例えば図14の7つの顔画像からなるフレームデータの集合であり、このフレームデータセット内の画像を音情報に合せて選択して出力すれば音声に同期した映像を合成することができる。音素片波形セットはテキストにあわせて音声を合成するときの各音の波形データの集合である。受信端末150では、受信したテキスト情報4とフレームデータセット153、音素片データセット156を用いて、音声・映像合成を行い出力する。

【0045】

図16は図15の受信端末150の一実施形態の構成図である。受信したテキスト情報4は通信IF60を介してメモリ166に格納される。メールを閲覧する前に、通信IF60を介して、フレームデータセット153及び音素片波形セット156を受信し、それぞれ、画像メモリ180と音素片メモリ161に格納する。ユーザの指示により、テキスト情報4、フレームデータセット153、音素片データセット156を用いて音声・映像合成を行うが、このときの処理は図12の処理とほぼ同じである。

【0046】

すなわち、音声変換部120、映像変換部125が必要なデータを決定し、データをアクセスする。データのアクセス先は図12の場合は音素片データベース132あるいは画像データベース128であったが、図16においては、図12の音素片データベース132の中の音声選択信号105により指定された音素片データセットのみが音素片メモリ161に格納されている。同様に、図12の画像データベース128の中の画像選択信号106により指定されたフレームデータセットのみが画像メモリ180に格納されている。画像の場合の例を以下に挙げる。

画像データベース128

選択信号 フレームデータ

1 CHILD0 CHILD1 CHILD2 CHILD3 CHILD4 CHILD5 CHILD6

2 MAN0 MAN1 MAN2 MAN3 MAN4 MAN5 MAN6

3 WOMAN0 WOMAN1 WOMAN2 WOMAN3 WOMAN4 WOMAN5 WOMAN6

画像メモリ 1 8 0

CHILD0 CHILD1 CHILD2 CHILD3 CHILD4 CHILD5 CHILD6

画像データベース 1 2 8 には 3 種類のフレームデータセットが格納されており、画像選択信号 1 0 6 によって選択される。例えば選択信号 = 1 の時には、CHILD0 から CHILD6 までの 6 フレームから成るフレームデータセットが合成に使用される。

【 0 0 4 7 】

一方、画像メモリ 1 8 0 では、すでに CHILD0 から CHILD6 までの 6 フレームから成るフレームデータセットが、画像データベース 1 5 2 からダウンロードされている。ダウンロード時には、例えば、画像データベース 1 5 2 の内容が画像データベース 1 2 9 と同じとすると、選択信号 1 5 1 として 1 を指定する。

【 0 0 4 8 】

このように、図 12 と同様に合成された音声 1 2 1 はスピーカ 7 8 から、映像 1 2 9 は表示デバイス 6 6 に出力される。また、ユーザの選択により、受信し、メモリ 1 6 6 に格納してあるテキスト情報自体をテキスト表示処理部 6 4 で、文字コードデータから文字ビットマップへの変換等を行った後に、表示デバイス 6 6 に出力することもできる。

【 0 0 4 9 】

テキスト情報の表示は、テキスト情報単独でも、映像情報の上に文字ビットマップをオーバーレイしても、画面の領域を分割し、一部に映像情報、別の部分にテキスト情報を表示しても構わない。また、テキスト情報の表示／非表示あるいは上記の表示形態はユーザが指定することができる。

【 0 0 5 0 】

上記本発明のマルチメディア配信システムの第 2 の実施形態では、音声・映像合成サーバが不要になり、配信サーバ 3 も、テキスト及び添付データを配信するだけの機能で済むため、構成が容易になる。また、配信サーバから受信端末への

トラフィックも一般的には第1の実施例に比べ少なくなり、低い通信料金で通信が可能になる。一方、受信端末150側は、音声・画像合成機能が端末内に必要になるため、装置規模は大きくなるものの、以下の利点がある。

【0051】

すなわち、受信者が自由な画像、音声を選択あるいは画像、音声による出力をしないことも選択できる点である。また、複数の音素片データセット及びフレームデータセットを受信者がダウンロードしておき、送信者候補リストとダウンロードした音声・画像の対応をあらかじめ指定しておくことにより、特定の送信者からのデータに対しては指定した音声・画像が出力させるようにする。また、音素片データセット及びフレームデータセットのデータフォーマット用いれば、利用者個人で音素片データセット及びフレームデータセットを作成し、作成したデータを用いて音声・映像合成を行うことができる。

【0052】

図17は本発明によるマルチメディア配信システムの第3の実施形態の構成図である。本実施形態では、第1の実施形態と同じ機能のサービス、すなわち、送信者が合成する音声、画像の種類を選択するサービスを実現する。

【0053】

図17において、送信端末200がテキスト情報を送信する前に、画像データベース152及び音素片データベース155に接続し、画像選択信号151及び音声選択信号154をそれぞれ送信することにより、フレームデータセット153及び音素片データセット156をダウンロードしておく。テキスト情報送信時には図18に示すように、先にダウンロードした画像情報311（フレームデータセット）と音素片情報312（音素片データセット）をテキスト情報51に付加し、さらに、これら画像情報311、音素片情報312が付加されていることを示す識別符号310を付加した情報を送信する。

【0054】

配信サーバ201、音声・映像合成サーバ204では送信端末200から送信されたテキスト情報、フレームデータセット、音素片データセットを使用して、音声・映像合成を行った後、テキスト情報と、音声・映像情報を受信端末5に送

信する。受信端末5は、図1の受信端末と同じである。

【0055】

図19は図17の送信端末200の一構成例の図である。送信端末200は、図8の送信端末100の合成音声・合成映像選択部110の代わりに、音素片メモリ202、画像メモリ204が置かれている。

ユーザは文字入力デバイス11、編集部13を用いて生成したテキスト情報14をメモリ15に格納する。テキスト情報14を送信する前に、通信IF201を用いて音素片データセット156及びフレームデータセット153をダウンロードして、それぞれ音素片メモリ202及び画像メモリ204に格納する。これらのダウンロードした情報は、図16の音素片メモリ161あるいは画像メモリ180に格納されている内容と同じである。テキスト情報16の送信時には、テキスト情報と16と、音素片データセット203及びフレームデータセット205を通信IF201を介して伝送路2に出力する。

【0056】

図20は配信サーバ201の構成図である。配信サーバ201の構成・動作は図11の配信サーバ101と、ほぼ同じ構成及び動作であるが、異なる点は音声・映像合成サーバ204に出力するデータが、配信サーバ101の場合では音声選択情報105と画像選択情報106が伝送されるのに対し、配信サーバ201では音素片データセット202、フレームデータセット203が伝送される点である。

【0057】

図21は音声・映像合成サーバ204の構成図である。音声・映像合成サーバ204の構成及び動作は図12の音声・映像合成サーバ103とほぼ同じである。異なる点は、音声・映像合成サーバ103では、音声選択信号105と画像選択信号106が入力され、それぞれ音素片データベース132、画像データベース128から合成に使用する音素片データセット、フレームデータセットが選択されるのに対して、音声・映像合成サーバ204の場合は、音素片データセット202とフレームデータセット210が入力され、それぞれ音素片メモリ132、画像メモリ220に格納され合成に使用される点である。

【 0 0 5 8 】

第 3 の実施形態の利点は、送信者が音声・画像データを選択する自由度が高くなる点である。すなわち、音素片・画像データベースが音声・映像合成サーバに含まれるような形態では、選択できる音声、画像の種類・料金等が音声・映像合成サーバの運営者によって制限される可能性があるが、第 3 の実施形態では、配信サーバの運営者、音声・映像合成サーバの運営者以外の複数の者が、音素片・画像データベースサーバを運営することが可能となり、市場競争原理により、利用できる音素片・画像の種類が増えたり、低料金でデータを利用することができたり、利用者に対する恩恵が多くなる。

【 0 0 5 9 】

さらに、一度ダウンロードした音素片・フレームデータセットを送信端末で記憶しておくことによって、常に同じ音声、画像を使用することができる。また、同一データフォーマットを使用することにより、例えば利用者個人の音声、画像を用いることもできる。

【 0 0 6 0 】

図 2 2 は本発明によるマルチメディア配信システムの第 4 の実施形態の構成図である。本実施形態では、第 1、3 の実施形態と同じ機能のサービス、すなわち、送信者が合成する音声、画像の種類を選択するサービスを実現する。

送信端末 2 0 0 は第 3 の実施形態の端末と同一のものであり、送信したデータも図 1 8 と同一である。配信サーバ 2 4 0 は受信したデータを指定した宛先に転送する機能のみを有する、いわゆる通常のメールサーバである。ここで、第 4 の実施形態が他の実施例と異なる点は、伝送路 4 で送信されるデータも、図 1 8 に示すデータと同じデータ構造、すなわち、テキスト情報 5 1 に識別符号 3 1 0、画像情報 3 1 1（フレームデータセット）と音素片情報 3 1 2 が付加された構造である点である。受信端末 2 5 0 は受信した、テキスト情報 5 1 に識別符号 3 1 0、画像情報 3 1 1（フレームデータセット）と音素片情報 3 1 2 を用いて音声・映像合成処理を端末内で行う。

【 0 0 6 1 】

図 2 3 は図 2 2 の受信端末 2 5 0 の構成図である。受信端末 2 5 0 の構造・動

作は、図 1 6 の受信端末 1 5 0 に類似しており、異なる点は、受信端末 1 5 0 が、音素片データセット 1 6 0、フレームデータセット 1 6 2 をそれぞれ、別の論理チャンネルから事前にダウンロードするのに対し、受信端末 2 5 0 ではこれら音素片データセット 1 6 0、フレームデータセット 1 6 2 が受信テキストデータ 1 6 5 に付加されているため、受信したデータをメモリ 1 6 6 にいったん格納した後に、音素片データセット 1 6 0、フレームデータセット 1 6 2 をメモリ 1 6 6 から抽出し、それぞれ音素片メモリ 1 6 1、画像メモリ 1 8 0 に格納する点である。

【 0 0 6 2 】

第 4 の実施形態の利点は、第 2 の実施形態に比べ、受信者が予め音素片、画像データをダウンロードする手間が不要な点、また、第 1 又は第 3 の実施形態と同じサービスを提供しながら、伝送路 4 上の伝送データ量を低減できる点である。

【 0 0 6 3 】

さらに、マルチメディア配信システムの第 5 の実施形態として、送信端末 1 0 0 から音声選択信号、画像選択信号を付加したテキスト情報を受信し、配信サーバが、画像データベース 1 5 2 と音素片データベース 1 5 5 からの音素片データセット、フレームデータセットをダウンロードし、受信したテキスト情報にこれら音素片データセット、フレームデータセット付加し、受信端末 2 5 0 に送信する構成である。第 5 の実施形態では、第 1、3、4 の実施形態と同じサービスを提供しながら、システム全体のトラフィックを最小にすることができる。

【 0 0 6 4 】

図 2 4 は本発明によるマルチメディア配信システムの第 6 の実施形態の構成図である。本実施形態と前述の 5 つの実施形態と異なる点は、変換処理の内容が、テキストから音声・顔画像ではなくメディア情報である点、すなわち映像符号から別方式あるいは別解像度（画像サイズ）の映像符号への変換である点である。送信端末 1 は従来知られている送信端末と同じように、送信端末 1 自らの中で撮影した映像を符号化し、音声等とともにテキスト情報に添付して信号 2 として配信サーバ 2 2 0 0 へ送信する。配信サーバ 2 2 0 0 では、他の実施形態と同様に端末データベースサーバ 1 0 7 に受信端末 5 の再生能力を問い合わせ、もし、受

信した信号 2 の符号化方式（例えば映像符号化方式）が問い合わせた再生可能な方式の中になければ、映像変換サーバ 2 2 0 2 に映像符号化方式の変換を要求する。

【 0 0 6 5 】

具体的には信号 2 中の映像符号の部分抽出し、抽出した映像符号 2 2 0 1 とその符号化方式 2 2 0 4 を出力し、また、受信端末 5 が再生可能な符号化方式と映像変換サーバ 2 2 0 2 の処理可能な符号化方式の中の共通方式の中から選んだ方式 1 0 8 を通知する。ここで信号 2 の映像符号方式 2 2 0 4 は、信号 2 の中に明示的に方式名等を示してもよく、映像添付ファイル名等から間接的に示唆してもよい。

【 0 0 6 6 】

映像変換サーバ 2 2 0 2 では映像符号 2 2 0 1 を符号化方式 1 0 8 にて示される方式に変換して変換映像符号 2 2 0 3 として出力する。配信サーバ 2 2 0 0 は変換映像符号 2 2 0 3 を元の映像符号（映像符号 2 2 0 1）に該当する部分と置き換え、信号 4 として受信端末 5 へ送信する。

【 0 0 6 7 】

図 2 5 は図 2 4 の配信サーバ 2 2 0 0 の構成図である。基本的な構成、動作は図 1 1 の配信サーバ 1 0 1 と同じであるが、入力信号 2 に変換元となる映像符号が含まれている点、音声映像合成サーバ 1 0 3 の代わりに、映像変換サーバ 2 2 0 2 に対し、映像符号 2 2 0 1 と映像符号方式 2 2 0 4 を送信し、変換された映像符号 2 2 0 3 を取得する点が異なる。また、映像符号 2 2 0 1 の符号化方式を取得するために、受信した情報 4 2 をコントロール部 2 1 0 3 に入力し、コントロール部 2 1 0 3 で、その符号化方式を解析する点が異なる。

【 0 0 6 8 】

図 2 6 は図 2 4 の映像変換サーバ 2 2 0 2 の構成図である。入力された映像符号 2 2 0 1 は映像デコーダ 2 2 1 0 に入力される。映像デコーダ 2 2 1 0 は複数の符号化方式を切り替えて処理する機能を有し、映像符号方式 2 2 0 4 で示された方式で映像を再生する。なお、映像符号方式 2 2 0 4 の代わりに映像符号 2 2 0 1 の中に記述された符号化方式情報を用いてもよい。再生された映像 2 2 1 1

はバッファ 2 2 1 2 へ格納された後、読み出されスケーリング部 2 2 1 4 へ入力される。スケーリング部 2 2 1 4 では画像サイズ、フレームレート、インターレース/プログレッシブ スキャン方式、色信号密度等の解像度を変換する。なお、画像サイズ等の変更がない場合はスケーリング部を迂回してもよい。また、予めスケーリング部 2 2 1 4 を省いてもよい。変換された映像はスイッチ 2 2 1 6 で選択された所定のエンコーダ 2 2 1 8 に供給される。エンコーダ 2 2 1 8 は映像符号化方式 1 0 8 で選択される。エンコードされた符号はスイッチ 2 2 1 9 を介して変換映像符号 2 2 0 3 として出力される。

【 0 0 6 9 】

第 6 の実施形態（図 2 4 から図 2 6）ではメディア情報として映像（動画像）から別方式・別解像度の映像への変換の例を示したが、以下のように変更してもよい。別解像度・同方式の映像、同解像度・別方式の映像、別ビットレートの映像、映像から映像の一部フレーム（静止画像）。また、メディア情報として音声・音響信号も同様な構成により別方式、別サンプリングレート、別帯域、別ビットレートへの変換が可能である。

【 0 0 7 0 】

変換前のメディア情報（入力メディア情報）と変換後のメディア情報（出力メディア情報）の組み合わせにより、異なる変換料金を送信者あるいは受信者に課金することができる。例えば下記のような例である。例では「→」の左側が入力メディア情報、右側が出力メディア情報、「：」の後が料金体系である。

例 1

高解像度動画像→低解像度動画像：出力動画像 1 秒につき 1 0 円

例 2

動画像→複数の静止画像：静止画像 1 枚につき 1 円

例 3

符号化された音声信号→別方式で符号化された音声信号：秒数にかかわらず 1 回
1 0 0 円

例 4

テキスト情報→符号化された音声＋顔画像動画像：変換基本料金 1 0 0 円＋テキ

スト情報の 1 文字 1 円

例 5

音声付動画像→別音声付動画像：解像度変換 1 回 1 0 0 円、フレームレート変換 1 回 2 0 円、ビットレート変換 1 回 3 0 円、音声符号化方式変換 1 回 1 0 0 円

上記例 1 では、例えば、図 2 6 のスケーリング部 2 2 1 4 が機能するごとに変換秒数を計測し、計測した秒数に応じて料金を計算することにより実現できる。

例 2 では、静止画像のエンコード回数すなわち出力枚数を計測することにより、また例 3 では音声符号変換処理の起動回数を計測することにより料金を計算することができる。例 4 では一連の変換処理を開始した時点で基本料金を課し、その後 1 文字分変換するごとに追加料金を基本料金に加算することにより実現できる。例 5 では、各変換部の動作／非動作に応じて課金料金を加算していくこともできるし、これらの処理を要求するコマンドを解析した時点で該当料金を計算して課金することもできる。なお、これら料金計算は、配信サーバ 2 2 0 1 内で計算課金しても、映像変換サーバ 2 2 0 2 内にて計算し、計算結果を配信サーバ 2 2 0 1 に通知し、配信サーバにて課金してもよい。

【 0 0 7 1 】

これら料金体系のうち、変換先の方式により、料金が変わるものは、変換先の方式が確定した時点すなわち、受信端末のメディア処理能力が判明した時点で、料金を計算し、送信端末に対し、計算した料金を提示し、送信端末側が料金を確認して了解指示を発行することにより始めて、変換料金の課金と、変換動作の実行を行うこともできる。

【 0 0 7 2 】

変換先の方式により複数の候補がある場合には、先の実施例では変換サーバが所定の優先順位に応じて 1 つの候補に決定する方法を説明した。しかし、複数の候補の料金が異なる場合には、それら複数の候補とそれぞれの変換料金を送信端末に通知し選択してもらうこともできる。なお、一定時間選択指示がない場合には自動的に定められた手順にて決定される候補をサーバ側にて選択し実行する変形例や、送信端末が事前に候補選択手順を定めて設定しておく方法、送信端末がメディア情報送信するのに付随して希望候補あるいは候補選択手順を指示する方

法などの変形例も本発明に含まれる。なお、候補選択手順の例としては、料金の最も安いものを指示する方法、変換後のパラメータ（解像度、フレームレート、ビットレート等）の限度を示しそれらの限度内に含まれるものを任意選択する方法、変換後のパラメータの希望値を示しそれに最も近い性能の候補を選択する方法等がある。

以上本発明の実施形態について説明したが、本発明は上記実施形態に限定されるものではない。例えば、以下の形態も、本発明に含まれる。

第1から第5の実施形態において、音素片データセットの音素片波形データ及びフレームデータセットの画像データは例えば、MPEG-4等の圧縮符号化法を用いて圧縮した形で伝送を行ってもよい。この場合、伝送データ量が少なくなるため、システム全体のトラフィック量の低減や、ユーザの通信料金の低減を図ることができる。

【0073】

第1から第5の実施形態では、テキストを伝送すると、テキストの内容に対応した音声と映像を出力することを前提としたが、出力は音声のみ、あるいは、映像のみであっても構わない。配信サーバの提供するサービスとして、音声のみ、あるいは映像のみのサービスを提供する場合は、提供しないサービスの処理部、サーバ等は不要となる。

【0074】

第1から第6の実施形態では、課金は配信サーバにおいて、送信するデータに対して行っているが、これはデータ量に応じた課金でも、送信端末と配信サーバとの接続時間に応じた課金でもよい。また、配信サーバと受信端末間の通信も、データ量に応じた課金でも、受信端末と配信サーバとの接続時間に応じた課金でも構わない。また、受信端末と配信サーバの間の通信料金を送信端末に課金することも可能である。音声合成の有無、あるいは映像合成の有無により追加の料金を上乗せして課金することもできる。

【0075】

なお、各実施形態では、受信端末は配信サーバから自動的にデータを送信されることを前提として説明したが、受信端末から配信サーバに対して接続を行い、

受信端末宛のデータの有無を配信サーバに対して問い合わせ、該当データがあった場合に、データを受信端末内に伝送することも本発明に含まれる。

【 0 0 7 6 】

図 1 5、1 7 の場合、画像データベースサーバ、音素片データベースサーバからのデータセットのダウンロードに対して課金することも可能である。

第 2、第 4、第 5 の実施形態の場合、受信端末にてダウンロードした音素片データセット及びフレームデータセットを、送信者を識別する符号と関連付けて保存し、以降、同じ送信者からのデータに対して保存していたデータセットを使用することもできる。

【 0 0 7 7 】

第 1 から第 6 の実施形態のいずれの場合も、送信端末と配信サーバ、配信サーバと受信端末の間は有線伝送でも、無線伝送でも構わない。また、回線交換でもパケット交換でも構わない。また、第 1、3 の実施形態において、配信サーバと音声・映像合成サーバとの間は有線、無線いずれでも構わない。また、回線交換、パケット交換いずれでも構わない。配信サーバと音声・映像合成サーバは同一の装置であっても構わない。

【 0 0 7 8 】

第 1 から第 5 の実施形態のいずれの場合も、合成音声の選択と、合成映像の選択は独立に（別々に）行う例を示したが、音声と映像をセットで選択する場合も本発明に包含される。この場合、配信サーバと音声・映像合成サーバの間の選択信号は 1 系統で済み、また、図 1 5、1 7 の画像データベースサーバ、音素片データベースサーバは 1 つのサーバに統一することができる。

【 0 0 7 9 】

図 1 2、2 1 において、エンコードした音声と、エンコードした画像は多重化して出力しているが、これらは独立した 2 つのデータとして多重せずに出力しても構わない。この時、それぞれのデータに再生時刻情報（タイムスタンプ、フレーム番号等）を付加することにより、再生時に音声と映像の同期を容易にとることができる。

【 0 0 8 0 】

図13、14において、音素片の種類と、その継続時間によって顔画像を選択し、呈示する例を用いたが、以下の変形例においても類似の効果が得られる。

図14の顔画像の数は7種類の例を示したが、それ以上の数の画像を用いてもよく。この場合はより自然な、あるいは多くの表情を呈示することができ、自然感が増す効果がある。

【0081】

音素片と顔画像との対応は必ずしていなくても類似の効果が得られる。例えば音声出力区間と特定の顔画像の対応、音声未出力区間と特定の顔画像の対応をとった場合も類似の効果が得られる。具体的には、音声出力区間は図14の顔画像0と顔画像1とを適当な間隔で交互に選択する例である。このとき、音声未出力区間（無音区間）では図13に示したように顔画像0と顔画像6を適当な間隔で呈示することによりまばたきの自然な感じをだすことができる。この変形例では、顔画像の数は第14図の顔画像0、1、6の3種類で済むために画像メモリの記憶容量、フレームデータセットの転送時間、画像データベースサーバの規模等を削減できる効果がある。

【0082】

音素片と顔画像とが対応していない別の変形例として、音声出力区間にはランダムな画像を呈示し、音声未出力区間（無音区間）では図13に示したように顔画像0と顔画像6を適当な間隔で呈示する方法である。この方法では、原画像シーケンスから、ランダムあるいは一定間隔でフレームをサンプリングし、サンプリングしたフレームをフレームデータセットとして使用することができるため、フレームデータセットを容易に作成することができる。

【0083】

上記、全ての実施例、変形例における処理は、ソフトウェア処理、ハードウェア処理あるいはソフトウェア・ハードウェアの混在処理のいずれでもよい。

【0084】

【発明の効果】

上述のように、本発明ではテキスト情報を基に音声、映像情報を合成して生成することにより、送信端末の処理量を低減し、端末の小型化、端末電池の長寿命

化を実現できる。

【図面の簡単な説明】

【図 1】

図 1 は本発明によるマルチメディア配信システムの第 1 の実施形態を示す構成ブロック図。

【図 2】

図 1 の音声映像再生能力情報 2 1 0 2 を取得する手順を示すフロー図

【図 3】

図 1 の端末 DB サーバ 1 0 7 における音声映像再生能力情報管理の一例を示す図。

【図 4】

配信サーバへ返信する端末能力送信フォーマットと音声映像再生能力情報の例を示す図。

【図 5】

図 1 の配信サーバ 1 0 1 における、音声映像再生能力情報の音声能力の処理フロー図。

【図 6】

図 5 の方式選択に優先順位を設けた選択法による処理フロー図。

【図 7】

本発明の配信システムに使用されるマルチメディア端末の構成図。

【図 8】

図 7 のマルチメディア端末 1 0 0 0 の送信機能のみを抽出した送信端末 1 0 0 の構成図。

【図 9】

図 8 の伝送路 2 で伝送される信号を示す図。

【図 1 0】

図 8 の合成音声・合成映像選択部 1 1 0 における音声・映像選択の画面図。

【図 1 1】

本発明による配信サーバの一実施形態の構成図。

【図 1 2】

本発明における音声・映像合成サーバの一実施形態の構成図。

【図 1 3】

図 1 2 における音声・映像合成の説明図。

【図 1 4】

図 1 2 における音声・映像合成の説明図。

【図 1 5】

本発明によるマルチメディア配信システムの第2の実施形態の構成図。

【図 1 6】

図 1 5 の受信端末 1 5 0 の一実施形態の構成図。

【図 1 7】

本発明によるマルチメディア配信システムの第 3 の実施形態の構成図。

【図 1 8】

図 1 7 の送信データの模式図。

【図 1 9】

図 1 7 の送信端末 2 0 0 の構成図。

【図 2 0】

図 1 7 の配信サーバ 2 0 1 の構成図。

【図 2 1】

図 1 7 の音声・画像合成サーバ 2 0 4 の構成図。

【図 2 2】

本発明によるマルチメディア配信システムの第 4 実施形態の構成図。

【図 2 3】

図 2 2 の受信端末 2 5 0 の構成図。

【図 2 4】

本発明によるマルチメディア配信システムの第 6 実施形態の構成図。

【図 2 5】

図 2 4 の配信サーバ 2 2 0 0 の構成図。

【図 2 6】

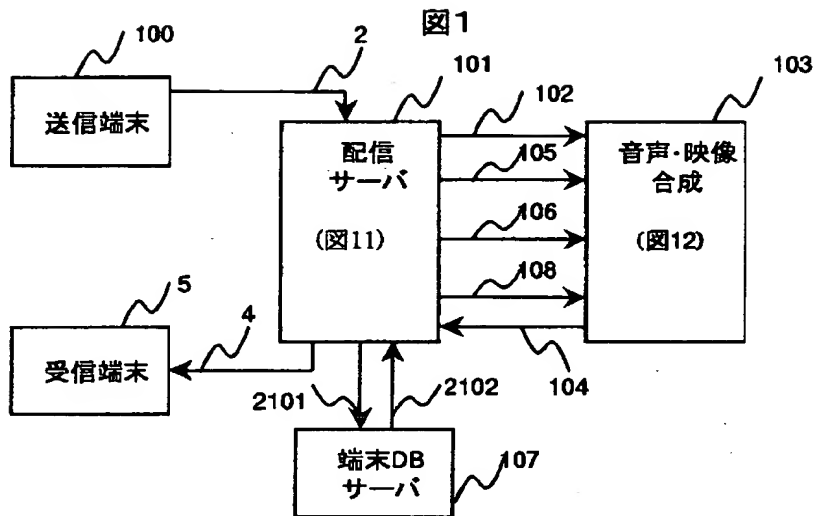
図 2 4 の映像変換サーバ 2 2 0 2 の構成図。

【符号の説明】

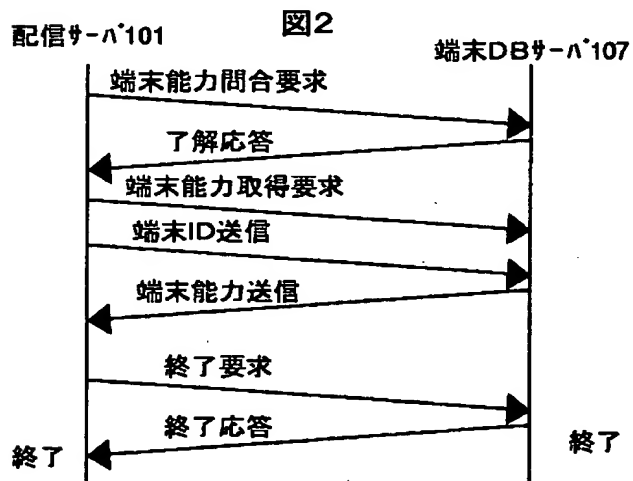
1 … 送信端末、3 … 配信サーバ、5 … 受信端末、1 0 0 … 送信端末、
1 0 3 … 音声・映像合成サーバ、1 0 7 … 端末データベースサーバ、
1 1 0 … 合成音声・映像選択部、1 2 5 … 映像変換部、
1 2 8 … 画像データベース、1 3 2 … 音素片データベース、1 3 4 … 音声変換部
、1 5 2 … 画像データベースサーバ、1 5 5 … 音素片データベースサーバ、
1 6 1 … 音素片メモリ、1 8 0 … 画像メモリ。

【書類名】 図面

【図 1】



【図 2】



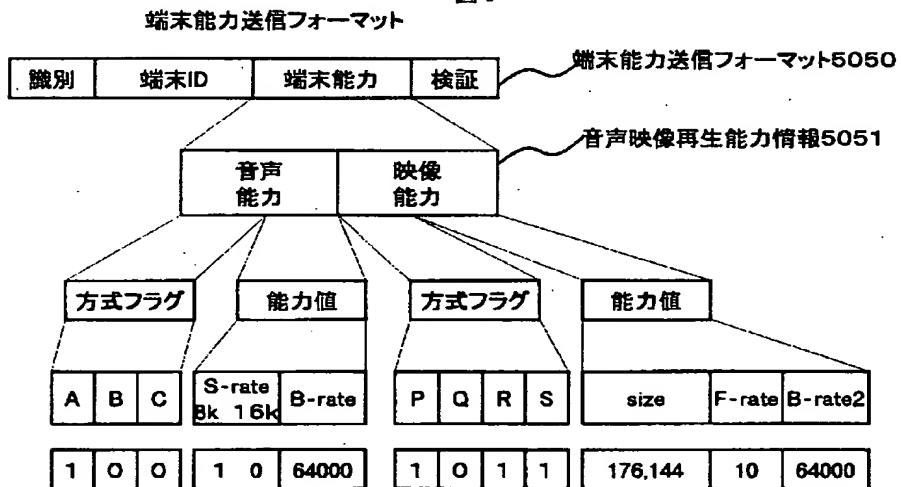
【図 3】

図3

端末ID1	端末能力1
端末ID2	端末能力2
端末ID3	端末能力3
...	...
端末ID N	端末能力 N

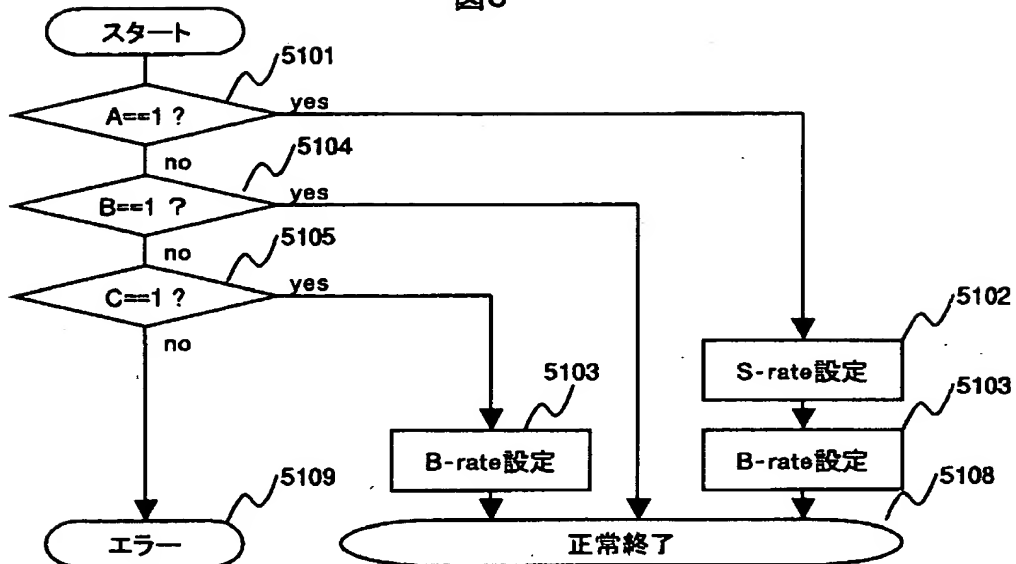
【図 4】

図4



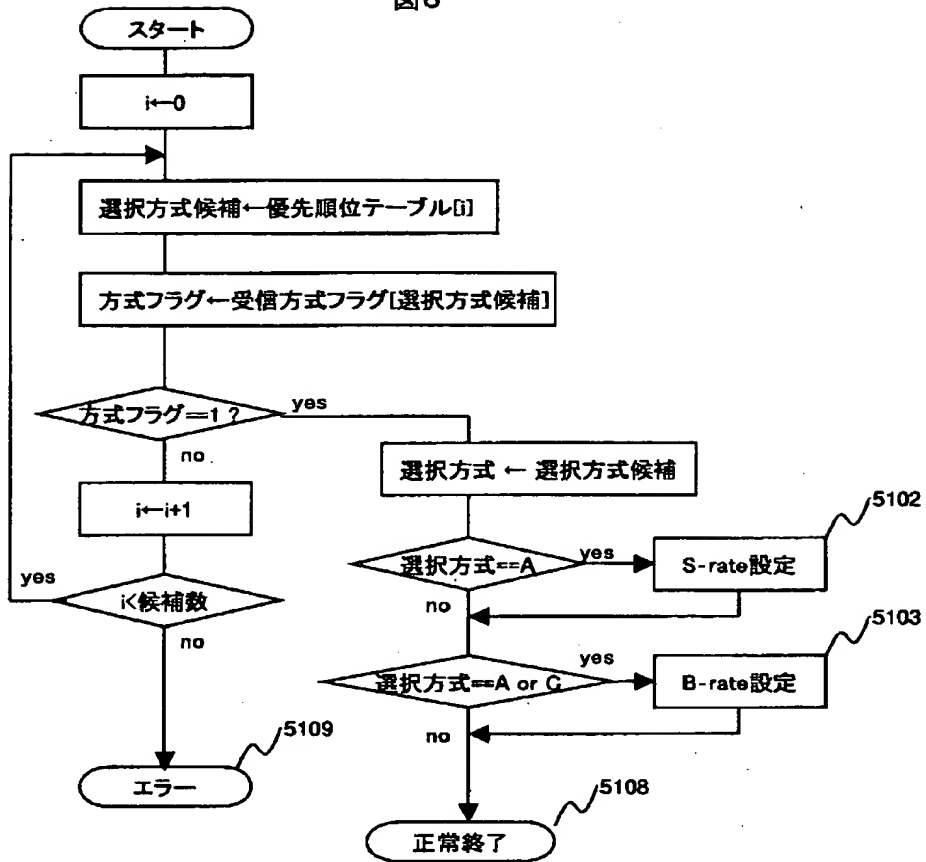
【図5】

図5

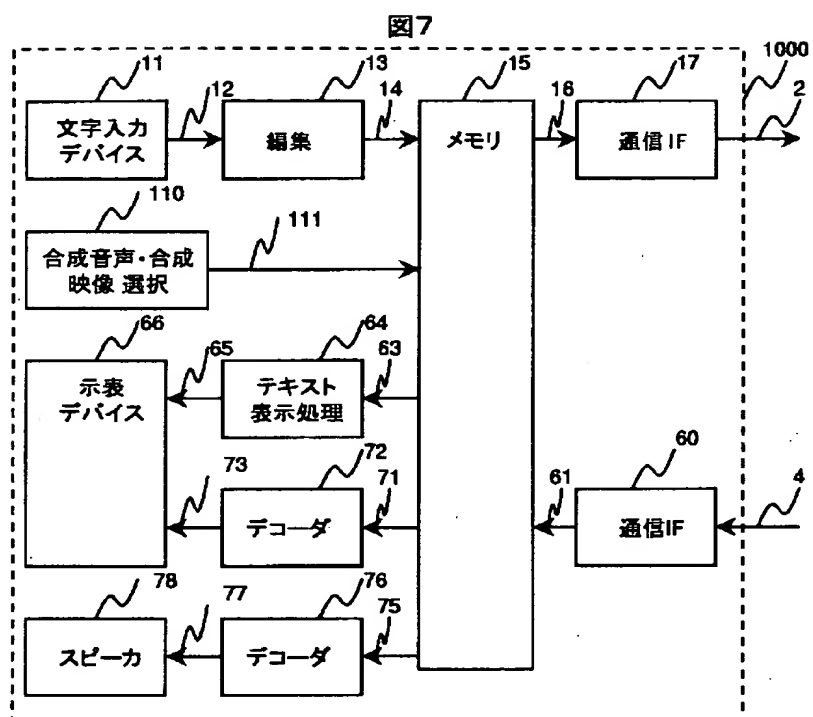


【図 6】

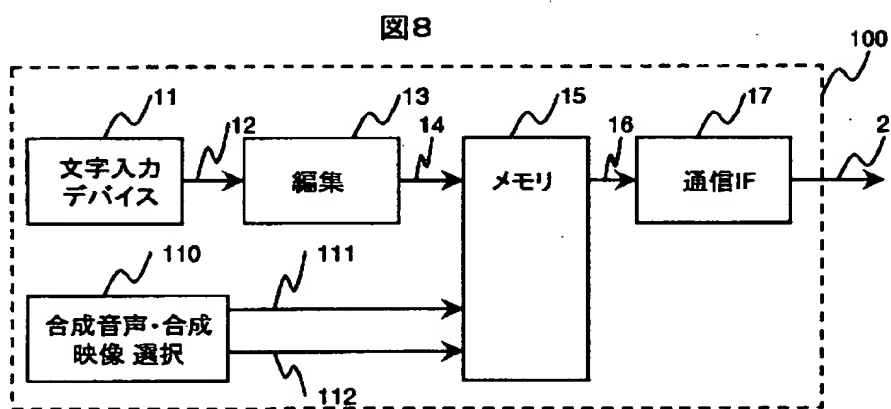
図 6



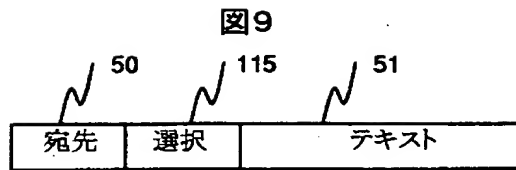
【图 7】



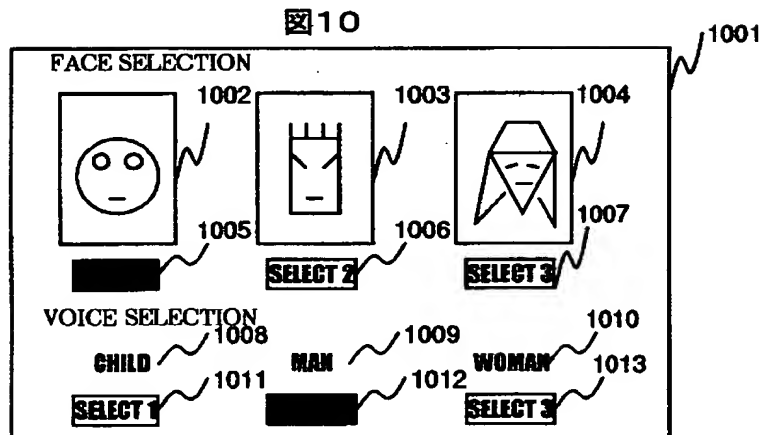
【图 8】



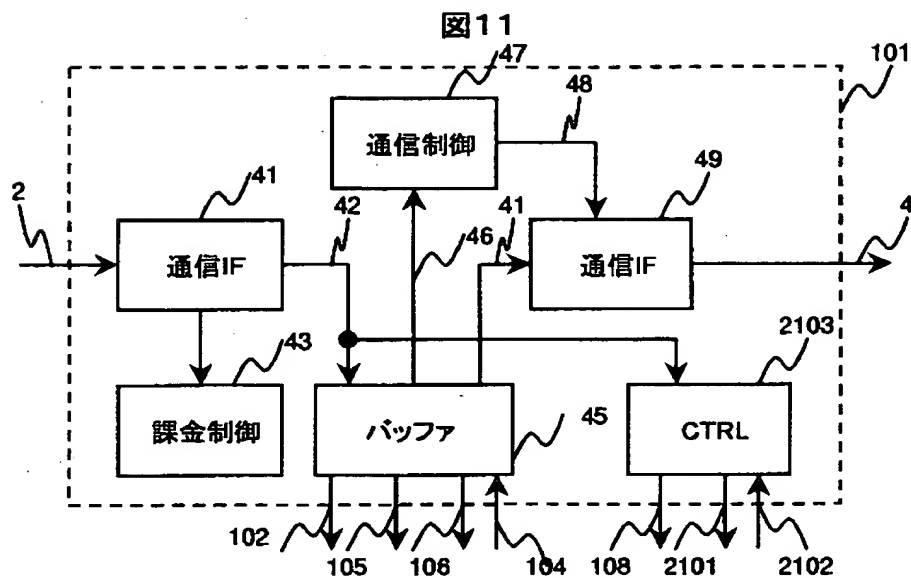
【図 9】



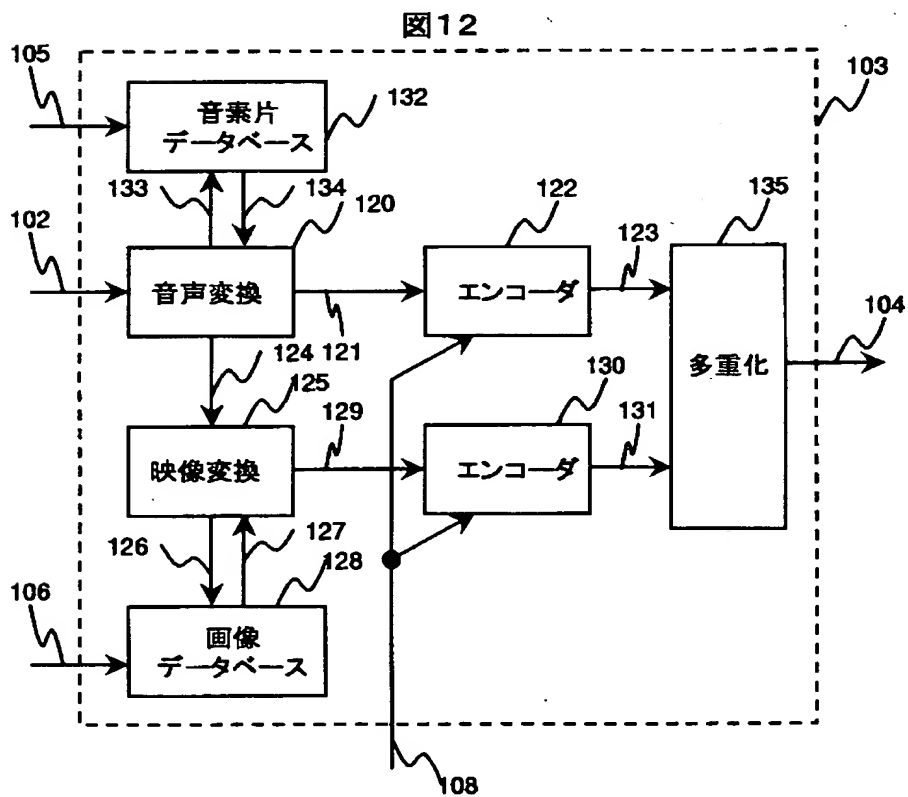
【図 1 0】



【図 1 1】



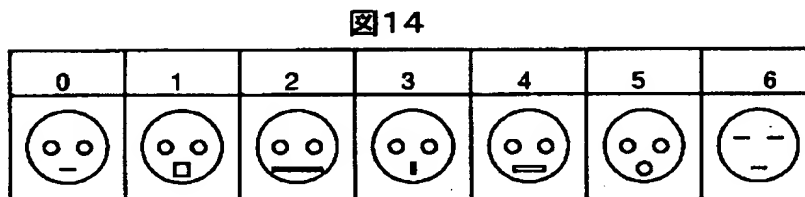
【図 1 2】



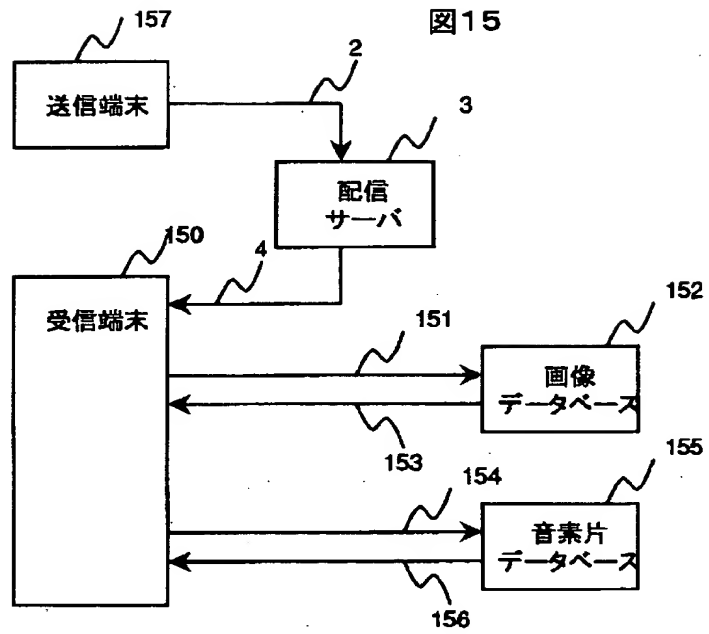
【図 13】



【図 14】

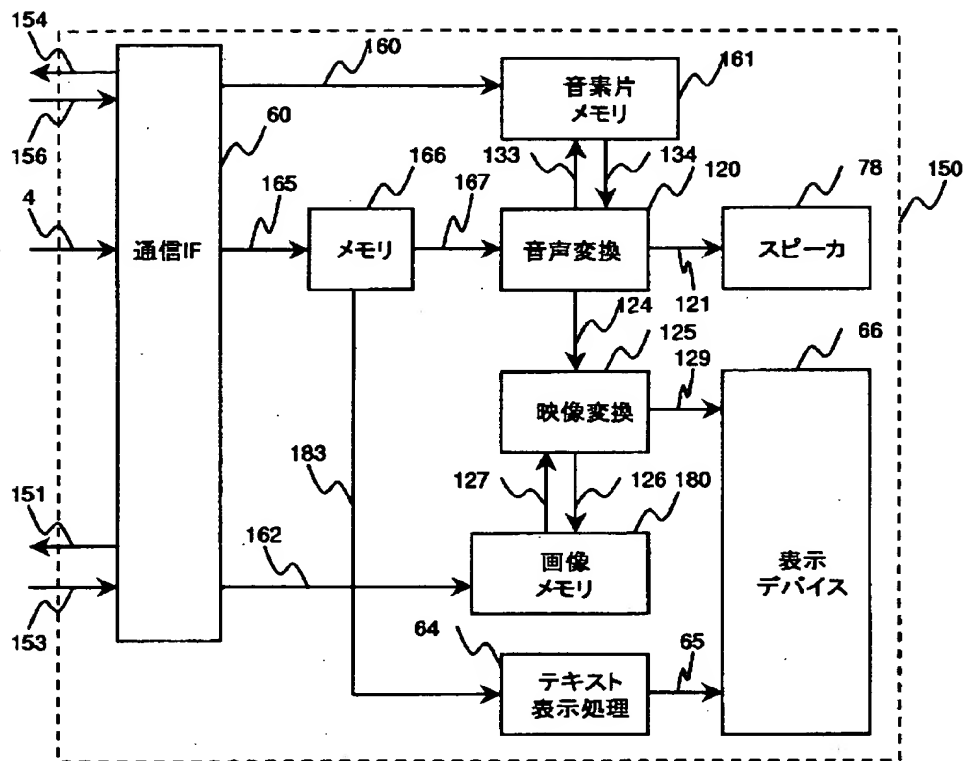


【図15】

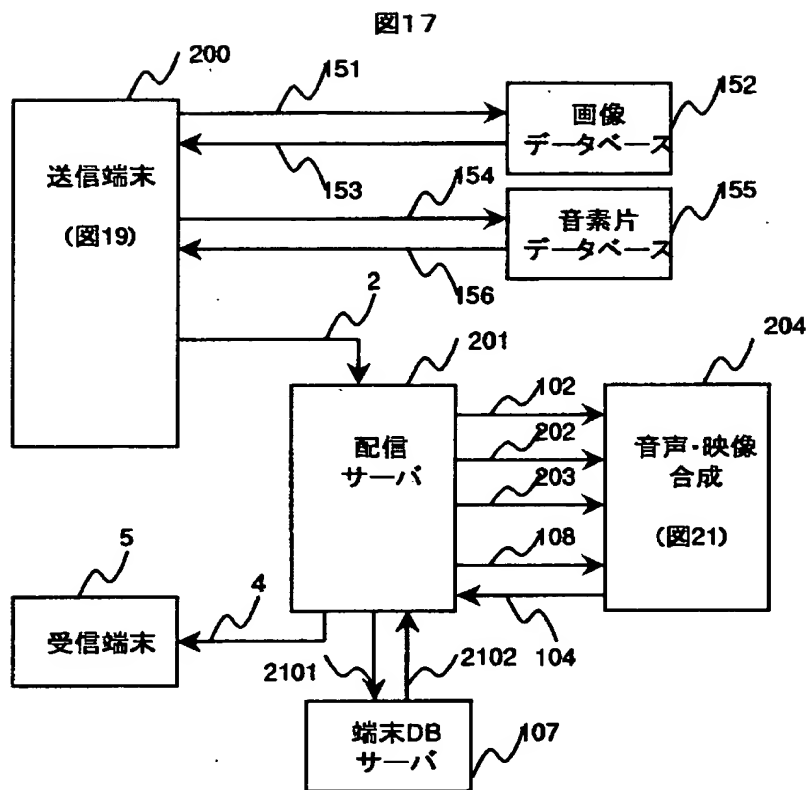


【図 16】

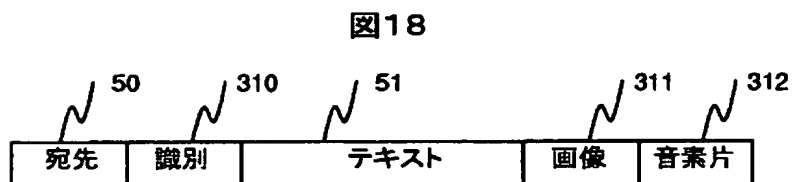
図16



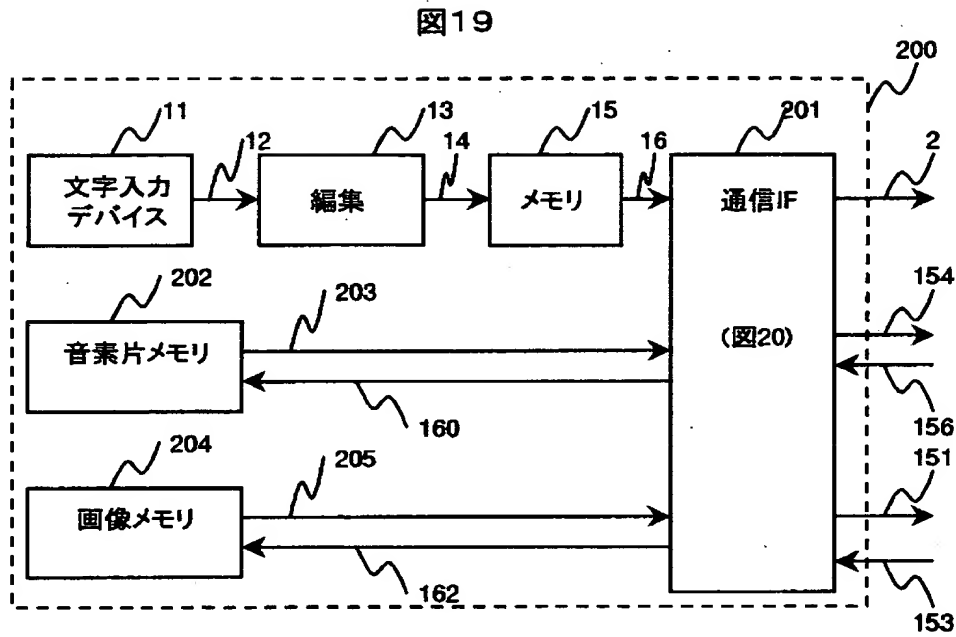
【図 17】



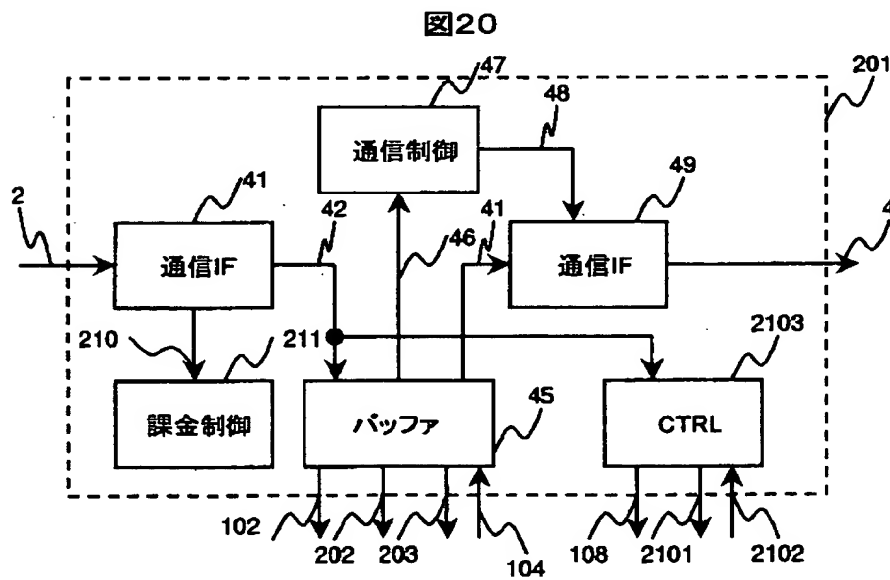
【図 18】



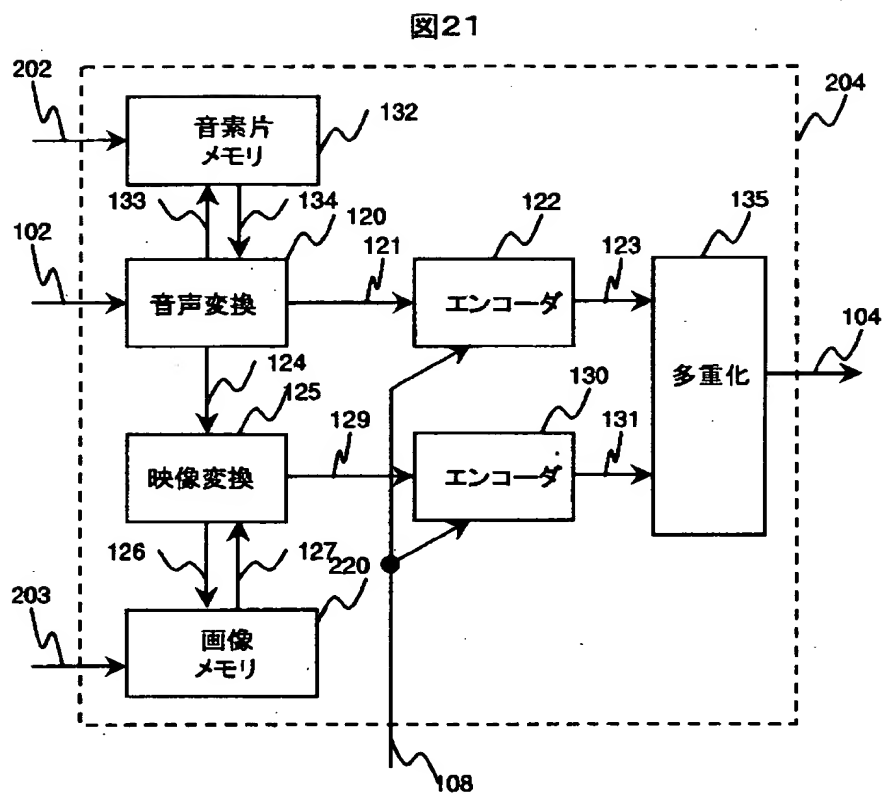
【図 1 9】



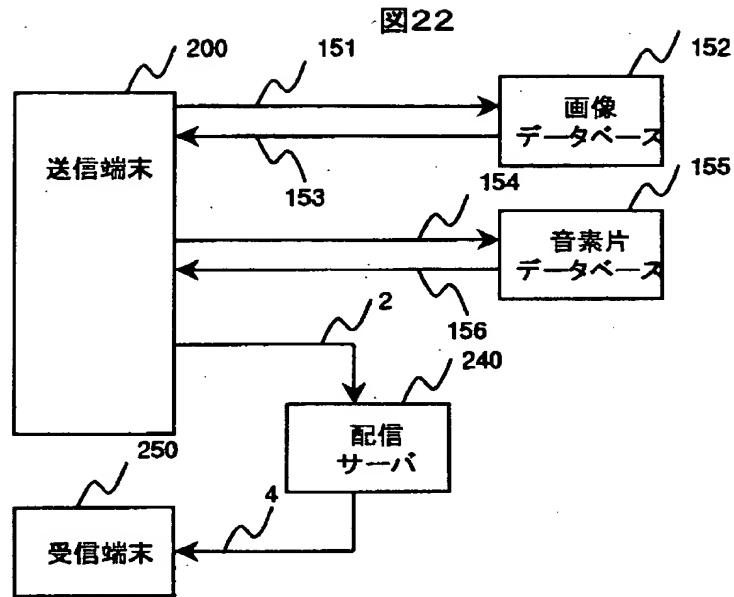
【図 2 0】



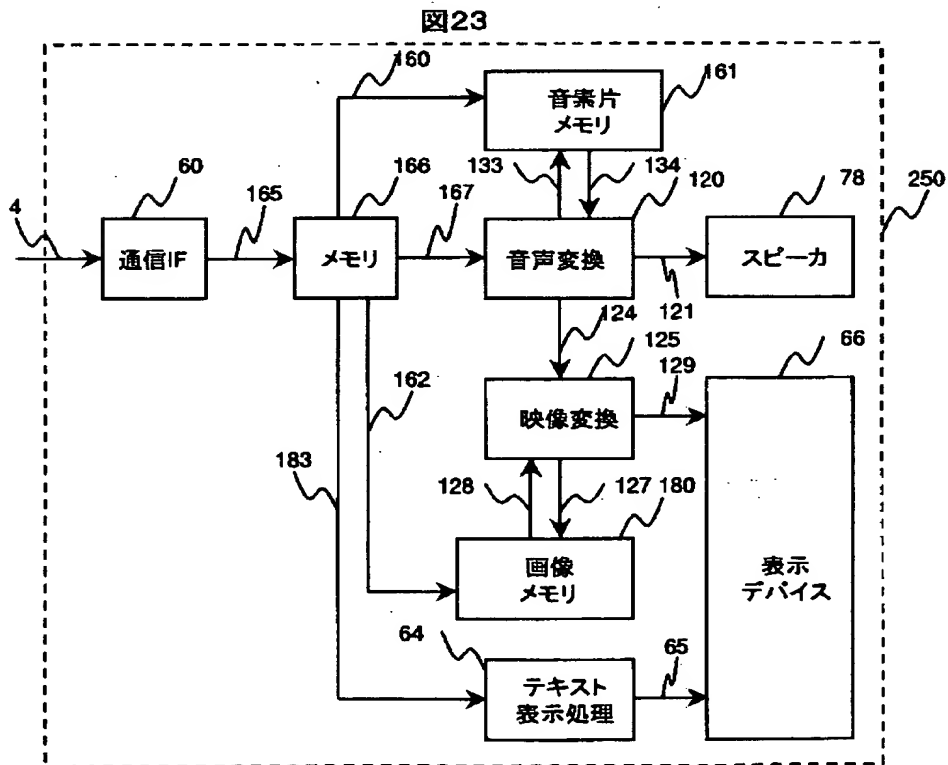
【図 21】



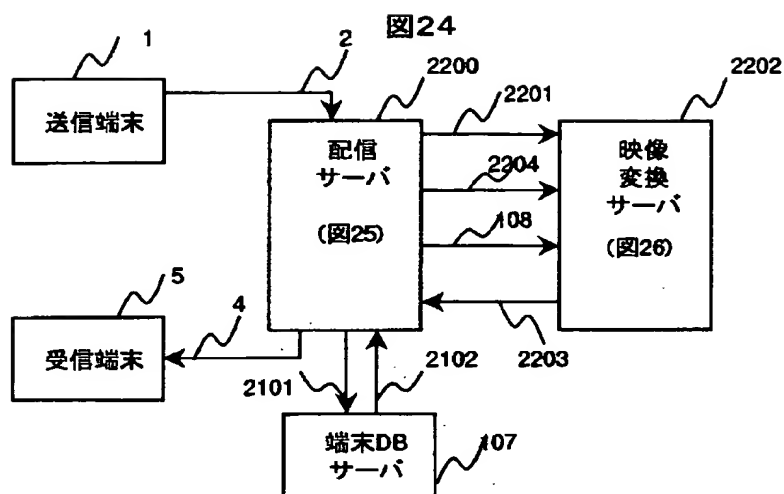
【図 22】



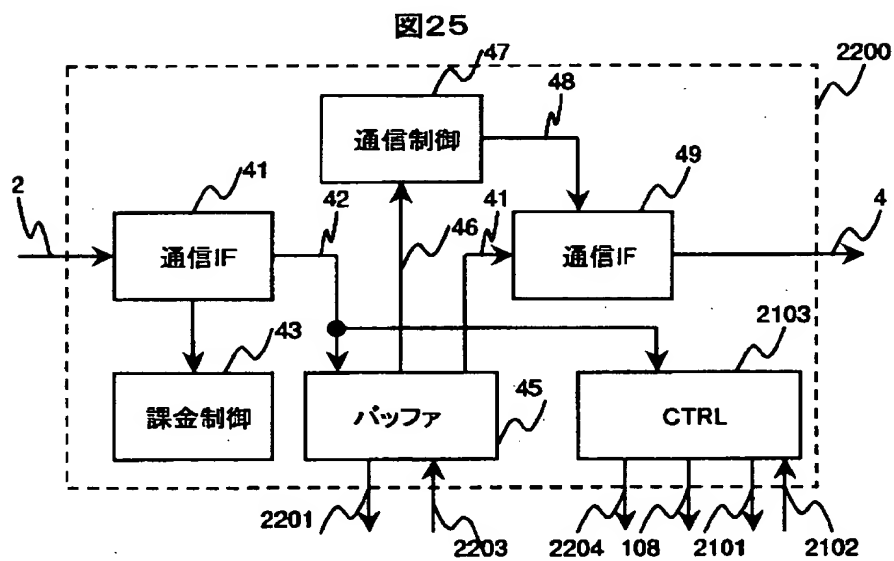
【図 23】



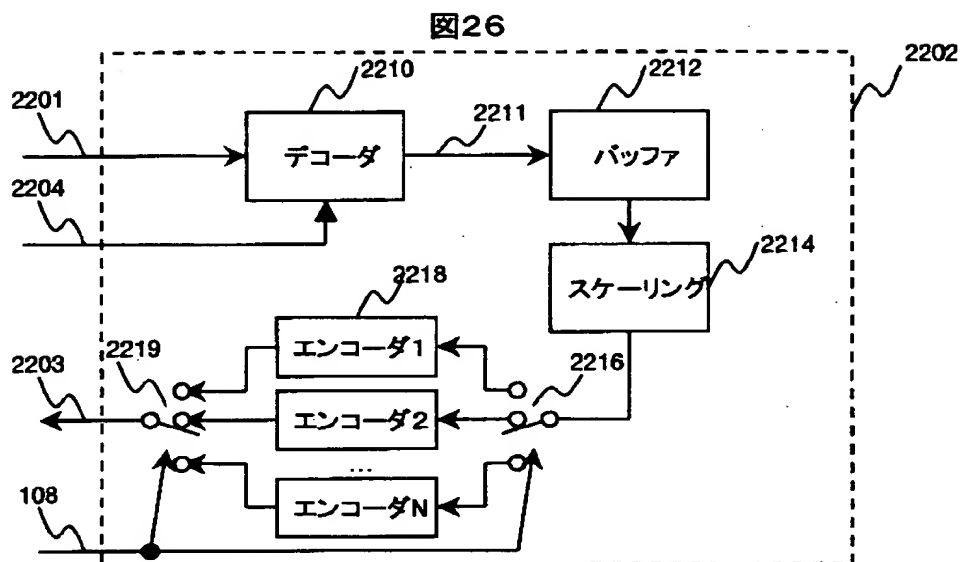
【図 24】



【図 25】



【図 26】



【書類名】 要約書

【要約】

【課題】 マルチメディア通信端末を経済的コストで、消費電力を少なく、かつ端末間の必要伝送容量を少なくする。

【解決手段】 送信、受信端末100、5間のマルチメディア通信データを中継するサーバを介してメディア情報を伝送・受信する配信システムにおいて、映像情報を配信サーバ101に付属の音声・映像合成サーバ103に予め格納しておき、通信時には格納していた映像情報を基に、受信端末5のメディア再生能力に応じた出力映像情報情報に変換する変換し、受信端末5に上記出力映像情報情報を送信する。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 1 0 8]

1. 変更年月日	1 9 9 0 年 8 月 3 1 日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台 4 丁目 6 番地
氏 名	株式会社日立製作所